

# Deciding Equivalence of Linear Tree-to-Word Transducers in Polynomial Time

Adrien Boiret<sup>\*1</sup> and Raphaela Palenta<sup>2</sup>

<sup>1</sup>CRISAL, University Lille 1, France

<sup>2</sup>Department of Informatics, Technical University of Munich,  
Germany

## Abstract

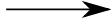
We show that the equivalence of deterministic linear top-down tree-to-word transducers is decidable in polynomial time. Linear tree-to-word transducers are non-copying but not necessarily order-preserving and can be used to express XML and other document transformations. The result is based on a partial normal form that provides a basic characterization of the languages produced by linear tree-to-word transducers.

## 1 Introduction

Tree transformations are widely used in functional programming and document processing. Tree transducers are a general model for transforming structured data like a database in a structured or even unstructured way. Consider the following internal representation of a client database that should be transformed to a table in HTML.

---

<sup>\*</sup>This work was partially supported by a grant from CPER Nord-Pas de Calais/FEDER DATA Advanced data science and technologies 2015-2020

<pre> { {   name: "Alexander"   surname: "Walker"   nickname: "Alex"   title: "Prof."   salutation: "Mr." } {   name: "Sara"   surname: "Tolin"   nickname: "Sara"   title: []   salutation: "Ms." } ... } </pre>		<pre> &lt;table&gt;   &lt;tr&gt;     &lt;th&gt; name &lt;\th&gt;     &lt;th&gt; surname &lt;\th&gt;     &lt;th&gt; nickname &lt;\th&gt;     &lt;th&gt; title &lt;\th&gt;     &lt;th&gt; salutation &lt;\th&gt; &lt;\tr&gt;   &lt;tr&gt;     &lt;td&gt; Alexander &lt;\td&gt;     &lt;td&gt; Walker &lt;\td&gt;     &lt;td&gt; Alex &lt;\td&gt;     &lt;td&gt; Prof. &lt;\td&gt;     &lt;td&gt; Mr. &lt;\td&gt; &lt;\tr&gt;   &lt;tr&gt;     &lt;td&gt; Sara &lt;\td&gt;     &lt;td&gt; Tolin &lt;\td&gt;     &lt;td&gt; Sara &lt;\td&gt;     &lt;td&gt; &lt;\td&gt;     &lt;td&gt; Ms. &lt;\td&gt; &lt;\tr&gt; &lt;\table&gt; </pre>
---	---	--

Top-down tree transducers can be seen as functional programs that transform trees from the root to the leaves with finite memory. Transformations where the output is not produced in a structured way or where, for example, the output is a string, can be modeled by tree-to-word transducers.

In this paper, we study deterministic linear tree-to-word transducers (LTWs), a subset of deterministic tree-to-word transducers that are non-copying, but not necessarily order-preserving. Processing the subtrees in an arbitrary order is important to avoid reordering of the internal data for different use cases. In the example of the client database the names may be needed in different formats, e.g.

```

<salutation> <name> <surname>
<surname>, <name>
<title> <surname>
<title> <surname>, <name>

```

The equivalence of unrestricted tree-to-word transducers was a long standing open problem that was recently shown to be decidable [14]. The algorithm by [14] provides an co-randomized polynomial algorithm for linear transducers. We show that the equivalence of LTWs is decidable in polynomial time and provide a partial normal form.

To decide equivalence of LTWs, we start in Section 3 by extending the methods used for sequential (linear and order-preserving) tree-to-word transducers (STWs), discussed in [15]. The equivalence for these transducers is decidable in polynomial time [15]. Moreover a normal form for sequential and linear tree-to-word transducers, computable in exponential time, is known [7, 1]. Two equivalent LTWs do not necessarily transform their trees in the same order. However, the differences that can occur are quite specific and characterized in [1]. We show how they can be identified. We use the notion of *earliest* states, inspired by the existing notion of earliest sequential transducers [7]. In this earliest form, two equivalent STWs can transform subtrees in different orders only if they fulfill specific properties pertaining to the periodicity of the words they create. Computing this normal form is exponential in complexity as the

number of states may increase exponentially. To avoid this size increase we do not compute these earliest transducers fully, but rather locally. This means we transform two LTWs with different orders to a *partial normal form* in polynomial time (see Section 4) where the order of their transformation of the different subtrees are the same. LTWs that transform the subtrees of the input in the same order can be reduced to sequential tree-to-word transducers as the input trees can be reordered according to the order in the transformation.

A short version of this paper will be published in the proceedings of the 20th International Conference on Developments in Language Theory (DLT 2016).

**Related Work.** Different other classes of transducers, such as tree-to-tree transducers [5], macro tree transducers [6] or nested-word-to-word transducers [15] have been studied. Many results for tree-to-tree transducers are known, e.g. deciding equivalence [12], minimization algorithms [12] and Gold-style learning algorithms [9]. In contrast, transformations where the output is not generated in a structured way like a tree are not that well understood. In macro-tree transducers, the decidability of equivalence is a well-known and long-standing question [2]. However, the equivalence of linear size increase macro-tree transducers that are equivalent to MSO definable transducers is decidable [3, 4].

## 2 Preliminaries

Let  $\Sigma$  be a ranked alphabet with  $\Sigma^{(n)}$  the symbols of rank  $n$ . Trees on  $\Sigma$  ( $\mathcal{T}_\Sigma$ ) are defined inductively: if  $f \in \Sigma^{(n)}$ , and  $t_1, \dots, t_n \in \mathcal{T}_\Sigma$ , then  $f(t_1, \dots, t_n) \in \mathcal{T}_\Sigma$  is a tree. Let  $\Delta$  be an alphabet. An element  $w \in \Delta^*$  is a word. For two words  $u, v$  we denote the concatenation of these two words by  $uv$ . The length of a word  $w$  is denoted by  $|w|$ . We call  $\varepsilon$  the empty word. We denote  $a^{-1}$  the inverse of a symbol  $a$  where  $aa^{-1} = a^{-1}a = \varepsilon$ . The inverse of a word  $w = u_1 \dots u_n$  is  $w^{-1} = u_n^{-1} \dots u_1^{-1}$ .

A *context-free grammar* (CFG) is defined as a tuple  $(\Delta, N, S, P)$ , where  $\Delta$  is the alphabet of  $G$ ,  $N$  is a finite set of *non-terminal symbols*,  $S \in N$  is the initial non-terminal of  $G$ ,  $P$  is a finite set of rules of form  $A \rightarrow w$ , where  $A \in N$  and  $w \in (\Delta \cup N)^*$ . A CFG is deterministic if each non-terminal has at most one rule.

We define the language  $L_G(A)$  of a non-terminal  $A$  recursively: if  $A \rightarrow u_0 A_1 u_1 \dots A_n u_n$  is a rule of  $P$ , with  $u_i$  words of  $\Delta^*$  and  $A_i$  non-terminals of  $N$ , and  $w_i$  a word of  $L_G(A_i)$ , then  $u_0 w_1 u_1 \dots w_n u_n$  is a word of  $L_G(A)$ . We define the context-free language  $L_G$  of a context-free grammar  $G$  as  $L_G(S)$ .

A *straight-line program* (SLP) is a deterministic CFG that produces exactly one word. The word produced by an SLP  $(\Delta, N, S, P)$  is called  $w_S$ .

We denote the *longest common prefix of all words* of a language  $L$  by  $\text{lcp}(L)$ . Its *longest common suffix* is  $\text{lcs}(L)$ .

A word  $u$  is said to be *periodic* of period  $w$  if  $w$  is the smallest word such that  $u \in w^*$ . A language  $L$  is said to be *periodic* of period  $w$  if  $w$  is the smallest word such that  $L \subseteq w^*$ .

A language  $L$  is *quasi-periodic* on the left (resp. on the right) of handle  $u$

and period  $w$  if  $w$  is the smallest word such that  $L \subseteq uw^*$  (resp. if  $L \subseteq w^*u$ ). A language is quasi-periodic if it is quasi-periodic on the right or left. If  $L$  is a singleton or empty, it is periodic of period  $\varepsilon$ . Iff  $L$  is periodic, it is quasi-periodic on the left and the right of handle  $\varepsilon$ . If  $L$  is quasi-periodic on the left (resp. right) then  $\text{lcp}(L)$  (resp.  $\text{lcs}(L)$ ) is the shortest word of  $L$ .

### 3 Linear Tree-to-Word Transducers

A *linear tree-to-word transducer* (LTW) is a tuple  $M = (\Sigma, \Delta, Q, \text{ax}, \delta)$  where

- $\Sigma$  is a ranked alphabet,
- $\Delta$  is an alphabet of output symbols,
- $Q$  is a finite set of states,
- the axiom  $\text{ax}$  is of the form  $u_0q(x)u_1$ , where  $q \in Q$  and  $u_0, u_1 \in \Delta^*$ ,
- $\delta$  is a set of rules of the form  $q, f \rightarrow u_0q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)})u_n$  where  $q, q_1, \dots, q_n \in Q$ ,  $f \in \Sigma$  of rank  $n$ ,  $u_0, \dots, u_n \in \Delta^*$  and  $\sigma$  is a permutation from  $\{1, \dots, n\}$  to  $\{1, \dots, n\}$ . There is at most one rule per pair  $q, f$ .

The partial function  $\llbracket M \rrbracket_q$  of a state  $q$  on an input tree  $f(t_1, \dots, t_n)$  is defined inductively as

- $u_0\llbracket M \rrbracket_{q_1}(t_{\sigma(1)}) \dots \llbracket M \rrbracket_{q_n}(t_{\sigma(n)})u_n$ , if  $q, f \rightarrow u_0q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)})u_n \in \delta$
- undefined, if  $q, f$  is not defined in  $\delta$ .

The partial function  $\llbracket M \rrbracket$  of an LTW  $M$  with axiom  $u_0q(x)u_1$  on an input tree  $t$  is defined as  $\llbracket M \rrbracket(t) = u_0\llbracket M \rrbracket_q(t)u_1$ .

Two LTWs  $M$  and  $M'$  are equivalent if  $\llbracket M \rrbracket = \llbracket M' \rrbracket$ .

A *sequential tree-to-word transducer* (STW) is an LTW where for each rule of the form  $q, f \rightarrow u_0q_1(x_{\sigma(1)})u_1 \dots q_n(x_{\sigma(n)})u_n$ ,  $\sigma$  is the identity on  $1 \dots n$ .

We define *accessibility* of states as the transitive and reflexive closure of appearance in a rule. This means state  $q$  is accessible from itself, and if  $q, f \rightarrow u_0q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)})u_n$ , and  $q$  is accessible from  $q'$ , then all states  $q_i$ ,  $1 \leq i \leq n$ , are accessible from  $q'$ .

We denote by  $\text{dom}(M)$  (resp.  $\text{dom}(q)$ ) the domain of an LTW  $M$  (resp. a state  $q$ ), i.e. all trees  $t \in \mathcal{T}_\Sigma$  such that  $\llbracket M \rrbracket(t)$  is defined (resp.  $\llbracket M \rrbracket_q(t)$ ). We only consider LTWs with non-empty domains and assume w.l.o.g. that no state  $q$  in an LTW has an empty domain by eliminating transitions using states with empty domain.

We denote by  $L_M$  (resp.  $L_q$ ) the range of  $\llbracket M \rrbracket$  (resp.  $\llbracket M \rrbracket_q$ ), i.e. the set of all images  $\llbracket M \rrbracket(t)$  (resp.  $\llbracket M \rrbracket_q(t)$ ). The languages  $L_M$  and  $L_q$  for each  $q \in Q$  are all context-free languages. We call a state  $q$  (*quasi-*)*periodic* if  $L_q$  is (quasi-)periodic.

Note that a word  $u$  in a rule of an LTW can be represented by an SLP without changing the semantics of the LTW. Therefore a set of SLPs can be added to the transducer and a word on the right-hand side of a rule can be represented by an SLPs. The decidability of equivalence of STWs in polynomial time still holds true with the use of SLPs. The advantage of SLPs is that they may compress the size of a word as the following example shows.

**Example 1.** We define an SLP  $G = (\Delta, N, S_0, P)$ , where  $N$  is a set  $\{S_0, \dots, S_n\}$ , the initial non-terminal is  $S_0$ , and  $P$  is the set of rules  $S_0 \rightarrow S_1 S_1$ ,  $S_1 \rightarrow S_2 S_2$ ,  $\dots$ ,  $S_{n-1} \rightarrow S_n S_n$ , and  $S_n \rightarrow a$ . This SLP produces the word  $a^{2^n}$ .  $G$  has  $n+1$  non-terminals and  $n+1$  rules. Thus,  $G$  produces a word that is exponential in the size of  $G$ .

The results of this paper require SLP compression to avoid exponential blow-up. SLPs are used to prevent exponential blow-up in [13], where morphism equivalence on context-free languages is decided in polynomial time.

The equivalence problem for sequential tree-to-word transducer can be reduced to the morphism equivalence problem for context-free languages [15]. This reduction relies on the fact that STWs transform their subtrees in the same order. As LTWs do not necessarily transform their subtrees in the same order the result cannot be applied on LTWs in general. However, if two LTWs transform their subtrees in the same order, then the same reduction can be applied. To formalize that two LTWs transform their subtrees in the same order we introduce the notion of state co-reachability. Two states  $q_1$  and  $q_2$  of LTWs  $M_1, M_2$ , respectively, are co-reachable if there is an input tree such that the two states are assigned to the same node of the input tree in the translations of  $M_1, M_2$ , respectively.

Two LTWs are *same-ordered* if for each pair of co-reachable states  $q_1, q_2$  and for each symbol  $f \in \Sigma$ , neither  $q_1$  nor  $q_2$  have a rule for  $f$ , or if  $q_1, f \rightarrow u_0 q'_1(x_{\sigma_1(1)}) \dots q'_n(x_{\sigma_1(n)}) u_n$  and  $q_2, f \rightarrow v_0 q''_1(x_{\sigma_2(1)}) \dots q''_n(x_{\sigma_2(n)}) v_n$  are rules of  $q_1$  and  $q_2$ , then  $\sigma_1 = \sigma_2$ .

If two LTWs are same-ordered the input trees can be reordered according to the order in the transformations. Therefore for each LTW a tree-to-tree transducer is constructed that transforms the input tree according to the transformation in the LTW. Then all permutations  $\sigma$  in the LTWs are replaced by the identity. Thus the LTWs can be handled as STWs and therefore the equivalence is decidable in polynomial time.

**Theorem 1.** *The equivalence of same-ordered LTWs is decidable in polynomial time.*

### 3.1 Linear Earliest Normal Form

In this section we introduce the two key properties that are used to build a normal form for linear tree-to-word transducers, namely the *earliest* and *erase-ordered* properties. The earliest property means that the output is produced as early as possible, i.e. the longest common prefix (resp. suffix) of  $L_q$  is produced in the rule in which  $q$  occurs, and as left as possible. The erase-ordered property means that all states that produce no output are ordered according to the input tree and pushed to the right in the rules.

An LTW is in *earliest form* if

- each state  $q$  is *earliest*, i.e.  $\text{lcp}(L_q) = \text{lcs}(L_q) = \varepsilon$ ,
- and for each rule  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$ , for each  $i, 1 \leq i \leq n$ ,  $\text{lcp}(L_{q_i} u_i) = \varepsilon$ .

In [1, Lemma 9] it is shown that for each LTW  $M$  an equivalent earliest LTW  $M'$  can be constructed in exponential time. Intuitively, if  $\text{lcp}(L_q) = v \neq \varepsilon$  (resp.  $\text{lcs}(L_q) = v \neq \varepsilon$ ) then  $q'$  is constructed with  $L_{q'} = v^{-1}L_q$  (resp.  $L_{q'} = L_qv^{-1}$ ) and  $q(x)$  is replaced by  $vq'(x)$  (resp.  $q'(x)v$ ). If  $\text{lcp}(L_q u) = v \neq \varepsilon$  and  $v$  is a prefix of  $u = vv'$  then we push  $v$  through  $L_q$  by constructing  $q'$  with  $L_{q'} = v^{-1}L_q v$  and replace  $q(x)u$  by  $vq'(x)v'$ .

Note that the construction to build the earliest form  $M'$  of an LTW  $M$  creates a same-ordered  $M'$ . Furthermore, if a state  $q$  of  $M$  and a state  $q'$  of  $M'$  are co-reachable, then  $q'$  is an “earliest” version of  $q$ , where some word  $u$  was pushed out of the production of  $q$  to make it earliest, and some word  $v$  was pushed through the production of  $q$  to ensure that the rules have the right property: there exists  $u, v \in \Delta^*$  such that for all  $t \in \text{dom}(q)$ ,  $\llbracket M' \rrbracket_{q'}(t) = v^{-1}u^{-1}\llbracket M \rrbracket_q(t)v$ .

**Theorem 2.** *For each LTW an equivalent same-ordered and earliest LTW can be constructed in exponential time.*

The exponential time complexity is caused by a potential exponential size increase in the number of states as it is shown in the following example.

We call a state  $q$  that produces only the empty word, i.e.  $L_q = \{\varepsilon\}$ , an *erasing state*. As erasing states do not change the transformation and can occur at any position in a rule we need to fix their position for a normal form.

An LTW  $M$  is *erase-ordered* if for each rule  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$  in  $M$ , if  $q_i$  is erasing then for all  $j \geq i$ ,  $q_j$  is erasing,  $\sigma(i) \leq \sigma(j)$  and  $u_j = \varepsilon$ .

We test whether  $L_q = \{\varepsilon\}$  in polynomial time and then reorder a rule according to the erase-ordered property. If an LTW is earliest it is still earliest after the reordering.

**Lemma 1** (extended from [1, Lemma 18]). *For each (earliest) LTW an equivalent (earliest) erase-ordered LTW can be constructed in polynomial time.*

**Example 2.** *Consider the rule  $q_0, f \rightarrow q_1(x_4)q_2(x_3)q_1(x_2)q_4(x_1)$  where  $q_2$  translates trees of the form  $f^n(g), n \geq 0$  to  $(abc)^n$ ,  $q_4$  translates trees of the form  $f^n(g), n \geq 0$  to  $(abc)^{2n}$ ,  $q_1$  translates trees of the form  $f^n(g), n \geq 0$  to  $\varepsilon$ . Thus the rule is not erase-ordered. We reorder the rule to the equivalent and erase-ordered rule  $q_0, f \rightarrow q_2(x_3)q_4(x_1)q_1(x_2)q_1(x_4)$ .*

If two equivalent LTWs are earliest and erase-ordered, then they are not necessarily same-ordered. For example, the rule  $q, f \rightarrow q_4(x_1)q_2(x_3)q_1(x_2)q_1(x_4)$  is equivalent to the rule in the above example but the two rules are not same-ordered. However, in earliest and erase-ordered LTWs, we can characterize the differences in the orders of equivalent rules: Just as two words  $u, v$  satisfy the equation  $uv = vu$  if and only if there is a word  $w$  such that  $u \in w^*$  and  $v \in w^*$ , the only way for equivalent earliest and erase-ordered LTWs to not be same-ordered is to switch periodic states.

**Theorem 3** ([1]). *Let  $M$  and  $M'$  be two equivalent erase-ordered and earliest LTWs and  $q, q'$  be two co-reachable states in  $M, M'$ , respectively. Let  $q, f \rightarrow u_0 q_1(x_{\sigma_1(1)}) \dots q_n(x_{\sigma_1(n)}) u_n$  and  $q', f \rightarrow v_0 q'_1(x_{\sigma_2(1)}) \dots q'_n(x_{\sigma_2(n)}) v_n$  be two rules for  $q, q'$ . Then*

- for  $k < l$  such that  $\sigma_1(k) = \sigma_2(l)$ , all  $q_i$ ,  $k \leq i \leq l$ , are periodic of the same period and all  $u_j = \varepsilon$ ,  $k \leq j < l$ ,
- for  $k, l$  such that  $\sigma_1(k) = \sigma_2(l)$ ,  $\llbracket M \rrbracket_{q_k} = \llbracket M' \rrbracket_{q'_l}$ .

As the subtrees that are not same-ordered in two equivalent earliest and erase-ordered states are periodic of the same period the order of these can be changed without changing the semantics. Therefore the order of these subtrees can be fixed such that equivalent earliest and erase-ordered LTWs are same-ordered. Then the equivalence is decidable in polynomial time, see Theorem 1. However, building the earliest form of an LTW is in exponential time.

To circumvent this difficulty, we will show that the first part of Theorem 3 still holds even on a *partial normal form*, where only quasi-periodic states are earliest and the longest common prefix of parts of rules  $q(x)u$  with  $L_q u$  being quasi-periodic is the empty word.

**Theorem 4.** *Let  $M$  and  $M'$  be two equivalent erase-ordered LTWs such that*

- *all quasi-periodic states  $q$  are earliest, i.e.  $\text{lcp}(q) = \text{lcs}(q) = \varepsilon$*
- *for each part  $q(x)u$  of a rule where  $L_q u$  is quasi-periodic,  $\text{lcp}(L_q u) = \varepsilon$*

*Let  $q, q'$  be two co-reachable states in  $M, M'$ , respectively and*

*$q, f \rightarrow u_0 q_1(x_{\sigma_1(1)}) \dots q_n(x_{\sigma_1(n)}) u_n$  and  $q', f \rightarrow v_0 q'_1(x_{\sigma_2(1)}) \dots q'_n(x_{\sigma_2(n)}) v_n$  be two rules for  $q, q'$ . Then for  $k < l$  such that  $\sigma_1(k) = \sigma_2(l)$ , all  $q_i$ ,  $k \leq i \leq l$ , are periodic of the same period and all  $u_j = \varepsilon$ ,  $k \leq j < l$ .*

## 4 Partial Normal Form

In this section we introduce a partial normal form for LTWs that does not suffer from the exponential blow-up of the earliest form. Inspired by Theorem 4, we wish to solve order differences by switching adjacent periodic states of the same period. Remember that the earliest form of a state  $q$  is constructed by removing the longest common prefix (suffix) of  $L_q$  to produce this prefix (suffix) earlier. It follows that all non-earliest states from which  $q$  can be constructed following the earliest form are quasi-periodic.

We show that building the earliest form of a quasi-periodic state or a part of a rule  $q(x)u$  with  $L_q u$  being quasi-periodic is in polynomial time. Therefore building the following partial normal form is in polynomial time.

**Definition 1.** *A linear tree-to-word transducer is in partial normal form if*

1. *all quasi-periodic states are earliest,*
2. *it is erase-ordered and*
3. *for each rule  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$  if  $L_{q_i} u_i L_{q_{i+1}}$  is quasi-periodic then  $q_i(x_{\sigma(i)}) u_i q_{i+1}(x_{\sigma(i+1)})$  is earliest and  $\sigma(i) < \sigma(i+1)$ .*

### 4.1 Eliminating Non-Earliest Quasi-Periodic States

In this part, we show a polynomial time algorithm to build an earliest form of a quasi-periodic state. From which an equivalent LTW can be constructed

in polynomial time such that any quasi-periodic state is earliest, i.e.  $\text{lcp}(L_q) = \text{lcs}(L_q) = \varepsilon$ . Additionally, we show that the presented algorithm can be adjusted to test if a state is quasi-periodic in polynomial time.

As quasi-periodicity on the left and on the right are symmetric properties we only consider quasi-periodic states of the form  $uw^*$  (quasi-periodic on the left). The proofs in the case  $w^*u$  are symmetric and therefore omitted here. In the end of this section we shortly discuss the introduced algorithms for the symmetric case  $w^*u$ .

To build the earliest form of a quasi-periodic state we use the property that each state accessible from a quasi-periodic state is as well quasi-periodic. However, the periods can be shifted as the following example shows.

**Example 3.** Consider states  $q$ ,  $q_1$  and  $q_2$  with rules  $q, f \rightarrow aq_1(x_1)c$ ,  $q_1, f \rightarrow aaq_2(x_1)ab$ ,  $q_2, f \rightarrow q_2(x_1)abc$ ,  $q_2, g \rightarrow abc$ . State  $q$  accepts trees of the form  $f^n(g)$ ,  $n \geq 2$ , and produces the language  $aaa(abc)^n$ , i.e.  $q$  is quasi-periodic of period  $abc$ . State  $q_1$  accepts trees of the form  $f^n(g)$ ,  $n \geq 1$ , and produces the language  $aa(abc)^nab$ , i.e.  $q_1$  is quasi-periodic of period  $cab$ . State  $q_2$  accepts trees of the form  $f^n(g)$ ,  $n \geq 0$  and produces the language  $(abc)^{n+1}$ , i.e.  $q_2$  is (quasi-)periodic of period  $abc$ .

We introduce two definitions to measure the shift of periods. We denote by  $\rho_n[u]$  the from right-to-left shifted word of  $u$  of shift  $n$ ,  $n \leq |u|$ , i.e.  $\rho_n[u] = u'^{-1}uu'$  where  $u'$  is the prefix of  $u$  of size  $n$ . If  $n \geq |u|$  then  $\rho_n[u] = \rho_m[u]$  with  $m = n \bmod |u|$ .

For two quasi-periodic states  $q_1, q_2$  of period  $u = u_1u_2$  and  $u' = u_2u_1$ , respectively, we denote the shift in their period by  $s(q_1, q_2) = |u_1|$ .

The size of the periods of a quasi-periodic state and the states accessible from this state can be computed from the size of the shortest words of the languages produced by these states.

**Lemma 2.** If  $q$  is quasi-periodic on the left with period  $w$ , and  $q'$  accessible from  $q$ , then  $q'$  is quasi-periodic with period  $\varepsilon$  or a shift of  $w$ . Moreover we can calculate the shift  $s(q, q')$  in polynomial time.

We now use these shifts to build, for a state  $q$  in  $M$  that is quasi-periodic on the left, a transducer  $M^q$  equivalent to  $M$  where each occurrence of  $q$  is replaced by its equivalent earliest form, i.e. a periodic state and the corresponding prefix.

**Algorithm 1.** Let  $q$  be a state in  $M$  that is quasi-periodic on the left.  $M^q$  starts with the same states, axiom, and rules as  $M$ .

- For each state  $p$  accessible from  $q$ , we add a copy  $p^e$  to  $M^q$ .
- For each rule  $p, f \rightarrow u_0q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)})u_n$  in  $M$  with  $p$  accessible from  $q$ , we add a rule  $p^e, f \rightarrow u_pq_1^e(x_{\sigma(1)})q_2^e(x_{\sigma(2)}) \dots q_n^e(x_{\sigma(n)})$  with  $u_p = \rho_{s(q,p)}[\text{lcp}(p)^{-1}u_0\text{lcp}(q_1) \dots \text{lcp}(q_n)u_n]$  in  $M^q$ .
- We delete state  $q$  in  $M^q$  and replace any occurrence of  $q(x)$  in a rule or the axiom of  $M^q$  by  $\text{lcp}(q)q^e(x)$ .



Note that  $\text{lcp}(p)^{-1}u_0\text{lcp}(q_1)\dots\text{lcp}(q_n)u_n$  is equivalent to deleting the prefix of size  $|\text{lcp}(p)|$  from the word  $u_0\text{lcp}(q_1)\dots\text{lcp}(q_n)u_n$ .

Intuitively, to build the earliest form of a state  $q$  that is quasi-periodic on the left we need to push all words and all longest common prefixes of states on the right-hand side of a rule of  $q$  to the left. Pushing a word to the left through a state needs to shift the language produced by this state. We explain the algorithm in detail on state  $q$  from Example 3.

**Example 4.** Remember that  $q$  produces the language  $aaa(abc)^n, n \geq 2$  and  $q_1, q_2$  accessible from  $q$  produce languages  $aa(abc)^nab, n \geq 1$  and  $(abc)^{n+1}, n \geq 0$ , respectively. Therefore  $\text{lcp}(q) = aaaabcbabc$ ,  $\text{lcp}(q_1) = aaabcb$  and  $\text{lcp}(q_2) = abc$ . We start with state  $q$ . As there is only one rule for  $q$  the longest common prefix of  $q$  and the longest common prefix of this rule are the same and therefore eliminated.

$$\begin{aligned} q^e, f &\rightarrow \rho_{s(q,q)}[\text{lcp}(q)^{-1}a\text{lcp}(q_1)c]q_1^e(x_1) \\ &\rightarrow \rho_{s(q,q)}[(aaaabcbabc)^{-1}aaaabcbabc]q_1^e(x_1) \\ &\rightarrow q_1^e(x_1) \end{aligned}$$

As there is only one rule for  $q_1$  the argumentation is the same and we get  $q_1^e, f \rightarrow q_2^e$ . For the rule  $q_2, f$  we calculate the longest common prefix of the right-hand side  $\text{lcp}(q_2)abc = abcbabc$  that is larger than the longest common prefix of  $q_2$ . Therefore we need to calculate the shift  $s(q, q_2) = s(q, q_1) + s(q_1, q_2) = |c| + |ab| = 3$  as  $q_1$  is accessible from  $q$  in rule  $q, f$  and  $q_2$  is accessible from  $q_1$  in rule  $q_1, f$ . This leads to the following rule.

$$\begin{aligned} q_2^e, f &\rightarrow \rho_{s(q,q_2)}[\text{lcp}(q_2)^{-1}\text{lcp}(q_2)abc]q_2^e(x_1) \\ &\rightarrow \rho_3[(abc)^{-1}abcbabc]q_2^e(x_1) \\ &\rightarrow abcbq_2^e(x_1) \end{aligned}$$

As the longest common prefix of  $q_2$  is the same as the longest common prefix of the right-hand side of rule  $q_2, g$  we get  $q_2^e, g \rightarrow \varepsilon$ . The axiom of  $M^q$  is  $\text{lcp}(q)q^e(x_1) = aaaabcbabcq^e(x_1)$ .

**Lemma 3.** Let  $M$  be an LTW and  $q$  be a state in  $M$  that is quasi-periodic on the left. Let  $M^q$  be constructed by Algorithm 1 and  $p^e$  be a state in  $M^q$  accessible from  $q^e$ . Then  $M$  and  $M^q$  are equivalent and  $p^e$  is earliest.

To replace all quasi-periodic states by their equivalent earliest form we need to know which states are quasi-periodic. Algorithm 1 can be modified to test an arbitrary state for quasi-periodicity on the left in polynomial time. The only difference to Algorithm 1 is that we do not know how to compute  $\text{lcp}(p)$  in polynomial time and  $s(q, p)$  does not exist. We therefore substitute  $\text{lcp}(p)$  by some smallest word of  $L_p$  and we define a mock-shift  $s'(q, p)$  as follows

- $s'(q, q) = 0$  for all  $q$ ,
- if  $q, f \rightarrow u_0q_1(x_{\sigma(1)})\dots q_n(x_{\sigma(n)})u_n$ , we say  $s'(q, q_i) = |u_iw_{q_{i+1}}\dots w_{q_n}u_n|$ , where  $w_q$  is a shortest word of  $L_q$ ,
- if  $s'(q_1, q_2) = n$  and  $s'(q_2, q_3) = m$  then  $s'(q_1, q_3) = n + m$ .

If several definitions of  $s'(q, p)$  exist, we use the smallest. If  $p$  is accessible from a quasi-periodic  $q$ , then  $s'(q, p) = s(q, p)$ .

**Algorithm 2.** Let  $M = (\Sigma, \Delta, Q, ax, \delta)$  be an LTW and  $q$  be a state in  $M$ . We build an LTW  $T^q$  as follows.

- For each state  $p$  accessible from  $q$ , we add a copy  $p^e$  to  $T^q$ .
- The axiom is  $w_q q^e(x)$  where  $w_q$  is a shortest word of  $L_q$ .
- For each rule  $p, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$  in  $M$  with  $p$  accessible from  $q$ , we add a rule

$$p^e, f \rightarrow u_p q_1^e(x_{\sigma(1)}) q_2^e(x_{\sigma(2)}) \dots q_n^e(x_{\sigma(n)})$$

in  $T^q$ , where  $u_p$  is constructed as follows.

- We define  $u = u_0 w_1 \dots w_n u_n$ , where  $w_i$  is a shortest word of  $L_{q_i}$ .
- Then we remove from  $u$  its prefix of size  $|w'|$ , where  $w'$  is a shortest word of  $L_p$ . We obtain a word  $u'$ .
- Finally, we set  $u_p = \rho_{s'(q,p)}[u']$ .

As the construction of Algorithms 1 and 2 are the same if the state  $q$  is quasi-periodic,  $\llbracket M \rrbracket_q$  and  $\llbracket T^q \rrbracket$  are equivalent if  $q$  is quasi-periodic. Moreover,  $q$  is quasi-periodic if  $\llbracket M \rrbracket_q$  and  $\llbracket T^q \rrbracket$  are equivalent.

**Lemma 4.** Let  $q$  be a state of an LTW  $M$  and  $T^q$  be constructed by Algorithm 2. Then  $M$  and  $T^q$  are same-ordered and  $q$  is quasi-periodic on the left if and only if  $\llbracket M \rrbracket_q = \llbracket T^q \rrbracket$  and  $q^e$  is periodic.

As  $M$  and  $T^q$  are same-ordered we can test the equivalence in polynomial time, cf. Theorem 1. Moreover testing a CFG for periodicity is in polynomial time and therefore testing a state for quasi-periodicity is in polynomial time.

Algorithm 2 can be applied to a part  $q(x)u$  of a rule to test  $L_q u$  for quasi-periodicity on the left. In this case for each rule  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$  a rule  $\hat{q}, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n u$  is added to  $M$  and each occurrence of the part  $q(x)u$  in a rule of  $M$  is replaced by  $\hat{q}(x)$ . We then apply the above algorithm to  $\hat{q}$  and test  $\llbracket M \rrbracket_{\hat{q}}$  and  $\llbracket T^{\hat{q}} \rrbracket$  for equivalence and  $\hat{q}^e$  for periodicity.

**Example 5.** Let  $q$  be a state with the rules  $q, f \rightarrow bcaq(x_1)$ ,  $q, g \rightarrow \varepsilon$ . Thus,  $q$  transforms trees of the form  $f^n(g)$ ,  $n \geq 0$  to  $(bca)^n$ . We use Algorithm 2 to test  $L_q bc$  for quasi-periodicity on the left. As explained above we introduce a state  $\hat{q}$  with the rules  $\hat{q}, f \rightarrow bca\hat{q}(x_1)$ ,  $\hat{q}, g \rightarrow bc$ . We now apply Algorithm 2 on  $\hat{q}$ . We build  $T^{\hat{q}} = \{\{f, g\}, \{a, b, c\}, \{\hat{q}^e\}, ax, \delta\}$  as follows. The axiom  $ax$  is  $bc\hat{q}^e(x_0)$  as the shortest word of  $L_{\hat{q}}$  is  $bc$ . For the rule  $\hat{q}, f$  we build  $u = bcabc$  as  $bc$  is the shortest word of  $\hat{q}$ . Then we obtain  $u' = abc$  and  $u_{\hat{q}} = \rho_{s'(\hat{q}, \hat{q})}[abc] = abc$ . Thus we get  $\hat{q}^e, f \rightarrow abc\hat{q}^e(x_1)$ . For the rule  $\hat{q}, g$  we build  $u = bc$  and obtain  $u' = \varepsilon$  as the shortest word of  $\hat{q}$  is  $bc$ . Thus we get  $\hat{q}^e, g \rightarrow \varepsilon$ .

$T^{\hat{q}}$  transforms trees of the form  $f^n(g)$  to  $bc(abc)^n$  and  $\hat{q}$  transforms trees of the form  $f^n(g)$  to  $(bca)^n bc$ . Thus, they are equivalent. Additionally  $\hat{q}^e$  is periodic with period  $abc$ . It follows that  $L_q bc$  is quasi-periodic.

We introduced algorithms to test states for quasi-periodicity on the left and to build the earliest form for such states. These two algorithms can be adapted for states that are quasi-periodic on the right. There are two main differences. First, as the handle is on the right the shortest word of a language  $L$  that is quasi-periodic on the right is  $\text{lcs}(L)$ . Second, instead of pushing words through a

periodic language to the left we need to push words through a periodic language to the right.

Hence, we can test each state  $q$  of an LTW  $M$  for quasi-periodicity on the left and right. If the state is quasi-periodic we replace  $q$  by its earliest form. Algorithm 1 and 2 run in polynomial time if SLPs are used. This is crucial as the shortest word of a CFG can be of exponential size, cf. Example 1. However, the operations that are needed in the algorithms, namely constructing the shortest word of a CFG and removing the prefix or suffix of a word, are in polynomial time using SLPs, cf. [11].

**Theorem 5.** *Let  $M$  be an LTW. Then an equivalent LTW  $M'$  where all quasi-periodic states are earliest can be constructed in polynomial time.*

## 4.2 Switching Periodic States

In this part we obtain the partial normal form by ordering periodic states of an erase-ordered transducer where all quasi-periodic states are earliest. Ordering means that if the order of the subtrees in the translation can differ, we choose the one similar to the input, i.e. if  $q(x_3)q'(x_1)$  and  $q'(x_1)q(x_3)$  are equivalent, we choose the second order. We already showed how we can build a transducer where each quasi-periodic state is earliest and therefore periodic. However, we need to make parts of rules earliest such that periodic states can be switched as the following example shows.

**Example 6.** *Consider the rule  $q, h \rightarrow q_1(x_2)bq_2(x_1)$  where  $q_1, q_2$  have the rules  $q_1, f \rightarrow bcabcaq_1(x)$ ,  $q_1, g \rightarrow \varepsilon$ ,  $q_2, f \rightarrow cabq_2(x)$ ,  $q_2, g \rightarrow \varepsilon$ . States  $q_1$  and  $q_2$  are earliest and periodic but not of the same period as a subword is produced in between. We replace the non-earliest and quasi-periodic part  $q_1(x_2)b$  by their earliest form. This leads to  $q, h \rightarrow bq_1^e(x_2)q_2(x_1)$  with  $q_1^e, f \rightarrow cabcabq_1^e(x)$ ,  $q_1^e, g \rightarrow \varepsilon$ . Hence,  $q_1^e$  and  $q_2$  are earliest and periodic of the same period and can be switched in the rule.*

To build the earliest form of a quasi-periodic part of a rule  $q(x)u$  each occurrence of this part is replaced by a state  $\hat{q}(x)$  and for each rule  $q, f \rightarrow u_0q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)})u_n$  a rule  $\hat{q}, f \rightarrow u_0q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)})u_n$  is added. Then we apply Algorithm 1 on  $\hat{q}$  to replace  $\hat{q}$  and therefore  $q(x)u$  by their earliest form. Iteratively this leads to the following theorem.

**Theorem 6.** *For each LTW  $M$  where all quasi-periodic states are earliest we can build in polynomial time an equivalent LTW  $M'$  such that each part  $q(x)u$  of a rule in  $M$  where  $L_q u$  is quasi-periodic is earliest.*

In Theorem 4 we showed that order differences in equivalent erase-ordered LTWs where all quasi-periodic states are earliest and all parts of rules  $q(x)u$  are earliest are caused by adjacent periodic states. As these states are periodic of the same period and no words are produced in between these states can be reordered without changing the semantics of the LTWs.

**Lemma 5.** *Let  $M$  be an LTW such that*

- $M$  is erase-ordered,
- all quasi-periodic states in  $M$  are earliest and
- each  $q_i(x_{\sigma(i)})u_i$  in a rule of  $M$  that is quasi-periodic is earliest.

Then we can reorder adjacent periodic states  $q_i(x_{\sigma(i)})q_{i+1}(x_{\sigma(i+1)})$  of the same period in the rules of  $M$  such that  $\sigma(i) < \sigma(j)$  in polynomial time. The reordering does not change the transformation of  $M$ .

We showed before how to construct a transducer with the preconditions needed in Lemma 5 in polynomial time. Note that replacing a quasi-periodic state by its earliest form can break the erase-ordered property. Thus we need to replace all quasi-periodic states by its earliest form *before* building the erase-ordered form of a transducer. Then Lemma 5 is the last step to obtain the partial normal form for an LTW.

**Theorem 7.** *For each LTW we can construct an equivalent LTW that is in partial normal form in polynomial time.*

### 4.3 Testing Equivalence in Polynomial Time

It remains to show that the equivalence problem of LTWs in partial normal form is decidable in polynomial time. The key idea is that two equivalent LTWs in partial normal form are same-ordered.

Consider two equivalent LTWs  $M_1, M_2$  where all quasi-periodic states and all parts of rules  $q(x)u$  with  $L_q u$  is quasi-periodic are earliest. In Theorem 4 we showed if the orders  $\sigma_1, \sigma_2$  of two co-reachable states  $q_1, q_2$  of  $M_1, M_2$ , respectively, for the same input differ then the states causing this order differences are periodic with the same period. The partial normal form solves this order differences such that the transducers are same-ordered.

**Lemma 6.** *If  $M$  and  $M'$  are equivalent and in partial normal form then they are same-ordered.*

As the equivalence of same-ordered LTWs is decidable in polynomial time (cf. Theorem 1) we conclude the following.

**Corollary 1.** *The equivalence problem for LTWs in partial normal form is decidable in polynomial time.*

To summarize, the following steps run in polynomial time and transform a LTW  $M$  into its partial normal form.

1. Test each state for quasi-periodicity. If it is quasi-periodic replace the state by its earliest form.
2. Build the equivalent erase-ordered transducer.
3. Test each part  $q_i(x_i)u_i$  in each rule from right to left for quasi-periodicity on the left. If it is quasi-periodic on the left replace the part by its earliest form.
4. Order adjacent periodic states of the same period according to the input order.

This leads to our main theorem.

**Theorem 8.** *The equivalence of LTWs is decidable in polynomial time.*

## 5 Conclusion

The equivalence problem for linear tree-to-word transducers can be decided in polynomial time. To prove this we used a reduction to the equivalence problem between sequential transducers [7], or more exactly, to an extension of this result to same-ordered transducers. This reduction hinges on two points. First, we showed that the only structural differences between two equivalent earliest linear transducers are caused by periodic languages which are interchangeable. The structural characteristic of periodic languages has been used in the normalization of STWs [7]. Second, we showed that if building a fully earliest transducer is potentially exponential, our reduction only requires quasi-periodic states to be earliest, which can be done in polynomial time. The use of the equivalence problem for morphisms on a CFG [13] and of properties on straight-line programs [10] is essential here as it was in [7, 8]. This leads to further research questions, starting with generalization of this result to all tree-to-words transducers. Furthermore, is it possible that these techniques can be used to decrease the complexity of some problems in other classes of transducer classes, such as top-down tree-to-tree transducers, where the equivalence problem is known to be between EXPTIME-Hard and NEXPTIME?

## References

- [1] Adrien Boiret. Normal form on linear tree-to-word transducers. In *Language and Automata Theory and Applications*, pages 439–451. Springer, 2016.
- [2] Joost Engelfriet. Some open question and recent results on tree transducers and tree languages. In *Formal Language Theory, Perspectives and Open Problems*, pages 241–286. Academic Press, 1980.
- [3] Joost Engelfriet and Sebastian Maneth. Macro tree translations of linear size increase are MSO definable. *SIAM Journal on Computing*, 32(4):950–1006, 2003.
- [4] Joost Engelfriet and Sebastian Maneth. The equivalence problem for deterministic MSO tree transducers is decidable. *Information Processing Letters*, 100(5):206–212, 2006.
- [5] Joost Engelfriet, Grzegorz Rozenberg, and Giora Slutzki. Tree transducers, L systems and two-way machines. In *Proceedings of the tenth annual ACM symposium on Theory of computing*, pages 66–74. ACM, 1978.

- [6] Joost Engelfriet and Heiko Vogler. Macro tree transducers. *Journal of Computer and System Sciences*, 31(1):71–146, 1985.
- [7] Grégoire Laurence, Aurélien Lemay, Joachim Niehren, Slawek Staworko, and Marc Tommasi. Normalization of sequential top-down tree-to-word transducers. In *Language and Automata Theory and Applications*, pages 354–365. Springer, 2011.
- [8] Grégoire Laurence, Aurélien Lemay, Joachim Niehren, Slawek Staworko, and Marc Tommasi. Learning sequential tree-to-word transducers. In *Language and Automata Theory and Applications*, pages 490–502, 2014.
- [9] Aurélien Lemay, Sebastian Maneth, and Joachim Niehren. A learning algorithm for top-down XML transformations. In *Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 285–296, 2010.
- [10] Markus Lohrey. Algorithmics on SLP-compressed strings: A survey. *Groups Complexity Cryptology*, 4(2):241–299, 2012.
- [11] Markus Lohrey. *The Compressed Word Problem for Groups*. Springer, 2014.
- [12] Sebastian Maneth and Helmut Seidl. Deciding equivalence of top-down XML transformations in polynomial time. In *Programming Language Technologies for XML*, pages 73–79, 2007.
- [13] Wojciech Plandowski. *The complexity of the morphism equivalence problem for context-free languages*. PhD thesis, Warsaw University, 1995.
- [14] Helmut Seidl, Sebastian Maneth, and Gregor Kemper. Equivalence of deterministic top-down tree-to-string transducers is decidable. In *IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 943–962, 2015.
- [15] Slawomir Staworko, Grégoire Laurence, Aurélien Lemay, and Joachim Niehren. Equivalence of deterministic nested word to word transducers. In *Fundamentals of Computation Theory*, pages 310–322. Springer, 2009.

## A Proof of Theorem 4

**Theorem.** *Let  $M$  and  $M'$  be two equivalent erase-ordered LTWs such that*

- *all quasi-periodic states  $q$  are earliest, i.e.  $\text{lcp}(q) = \text{lcs}(q) = \varepsilon$*
- *for each part  $q(x)u$  of a rule where  $L_q u$  is quasi-periodic,  $\text{lcp}(L_q u) = \varepsilon$*

*Let  $q, q'$  be two co-reachable states in  $M, M'$ , respectively and*

*$q, f \rightarrow u_0 q_1(x_{\sigma_1(1)}) \dots q_n(x_{\sigma_1(n)}) u_n$  and  $q', f \rightarrow v_0 q'_1(x_{\sigma_2(1)}) \dots q'_n(x_{\sigma_2(n)}) v_n$  be two rules for  $q, q'$ . Then for  $k < l$  such that  $\sigma_1(k) = \sigma_2(l)$ , all  $q_i, k \leq i \leq l$ , are periodic of the same period and all  $u_j = \varepsilon, k \leq j < l$ .*

*Proof.* Let  $M_e$  and  $M'_e$  be the equivalent earliest transducer of  $M$  and  $M'$ , respectively, such that  $M$  and  $M_e$  as well as  $M'$  and  $M'_e$  are same-ordered (cf. Theorem 2).

Suppose there exists co-reachable (and thus equivalent) states  $q^e$  and  $q'^e$  in  $M_e$  and  $M'_e$ , respectively, with rules

$$\begin{aligned} q^e, f &\rightarrow v_0 q_1^e(x_{\sigma(1)}) \dots q_n^e(x_{\sigma(n)}) v_n, \\ q'^e, f &\rightarrow v'_0 q_1'^e(x_{\sigma'(1)}) \dots q_n'^e(x_{\sigma'(n)}) v'_n \end{aligned}$$

such that  $\sigma \neq \sigma'$ .

Let  $i$  be the first index such that  $\sigma(i) \neq \sigma'(i)$ . Following Theorem 3, we have  $j, j'$  such that  $\sigma'(j') = \sigma(i)$  and  $\sigma(j) = \sigma'(i)$  and all  $q_l^e$ ,  $i \leq l \leq j$  are periodic with the same period.

Let  $q$  and  $q'$  be the states in  $M$  and  $M'$ , respectively, from which the co-reachable states  $q^e$  and  $q'^e$  were constructed with the earliest construction proposed by [7]. From the earliest construction it follows that  $q$  and  $q'$  are co-reachable. Since the construction preserves the rule structure, we have:

$$\begin{aligned} q, f &\rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n \\ q', f &\rightarrow u'_0 q_1'(x_{\sigma'(1)}) \dots q_n'(x_{\sigma'(n)}) u'_n \end{aligned}$$

The earliest construction gives us that for all  $l \in \{1, \dots, n\}$ ,  $\llbracket M_e \rrbracket_{q_l^e}(t) = v^{-1} u^{-1} \llbracket M \rrbracket_{q_l}(t) v$  for some  $u, v \in \Delta^*$ . This means that if  $q_l^e$  is periodic, then  $q_l$  is quasi-periodic in its non-earliest form. The same is true for all  $q_l'$ .

However, the first property we supposed of  $M$  and  $M'$  implies that all those  $q_l$  and  $q_l'$  that are quasi-periodic are not only quasi-periodic, but periodic. Consider a part of the rule  $q_i(x_{\sigma(i)}) u_i \dots q_j(x_{\sigma(j)})$  that is periodic in the earliest form and therefore quasi-periodic in the non-earliest form. The first condition gives us that  $q_i, \dots, q_j$  are periodic. However, then the words  $u_i, \dots, u_{j-1}$  are not necessarily empty. As the part  $q_i(x_{\sigma(i)}) u_i \dots q_j(x_{\sigma(j)})$  is quasi-periodic we know that each part  $q_k(x_{\sigma(k)}) u_k$ ,  $i \leq k < j$  is quasi-periodic. Then the second condition of this theorem guarantees that the parts  $q_k(x_{\sigma(k)}) u_k$ ,  $i \leq k < j$  are not only quasi-periodic, but periodic. From which it follows that the words  $u_i, \dots, u_{j-1}$  are empty. As the part  $q_i(x_{\sigma(i)}) u_i \dots q_j(x_{\sigma(j)})$  is periodic and  $u_i, \dots, u_{j-1}$  are empty we get that  $q_i, \dots, q_j$  are periodic of the same period. The same holds true for states of a part of the rule  $q'_i(x_{\sigma'(i)}) u'_i \dots q'_j(x_{\sigma'(j)})$  that is periodic in the earliest form.  $\square$

## B Proof of Lemma 2

**Lemma.** *If  $q$  is quasi-periodic on the left with period  $w$ , and  $q'$  accessible from  $q$ , then  $q'$  is quasi-periodic with period  $\varepsilon$  or a shift of  $w$ . Moreover we can calculate the shift  $s(q, q')$  in polynomial time.*

*Proof.* This is done as an iterative proof with the following elementary step: If  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$ , and  $q$  is quasi-periodic on the left with handle  $u$  and period  $w$ , then for all  $i$  between 1 and  $n$ ,  $q_i$  is quasi-periodic with period  $\varepsilon$  or a shift of  $w$ .

We pick  $v_j$  a smallest word produced by state  $q_j$ . We then have that for all  $t \in \text{dom}(q_i)$ ,  $u_0 v_1 \dots u_{i-1} \llbracket M \rrbracket_{q_i}(t) u_i \dots v_n u_n \in L_q$ . If we call  $u_l = u_0 v_0 \dots u_{l-1}$  and

$u_r = u_i \dots v_n u_n$ , we obtain that  $L_{q_i} \subseteq u_l^{-1} L_q u_r^{-1}$ . Since  $L_q \subseteq uw^*$ , we can say  $L_{q_i} \subseteq u_l^{-1} (uw^*) u_r^{-1}$ . It is a classical result of regular languages that  $(uw^*) u_r^{-1}$  is either empty, a singleton, or a quasi-periodic language of period  $u_r w u_r^{-1}$ . By further removing a prefix to this language the period does not change. Hence, we get that  $u_l^{-1} (uw^*) u_r^{-1}$  is also either empty, a singleton, or a quasi-periodic language of period  $u_r w u_r^{-1}$ . This means that  $q_i$  is quasi-periodic, of period  $\varepsilon$ , or  $u_r w u_r^{-1}$ , which is a shift of  $q$ . The size of  $u_r$  can easily be computed from the sizes of the minimal productions of states  $q_j$ . We build the CFG for  $L_{q_j}$ . Then, finding the smallest production of  $q_j$  and their size is finding the smallest word of  $L_{q_j}$  and their size, which is a polynomial problem on CFG.

To show that the shifts of the periods can be calculated in polynomial time we show that shifts are additive in nature: If  $q_1$  has period  $w$ ,  $q_1$  and  $q_2$  are of shifted period, and  $q_2$  and  $q_3$  are of shifted period, then  $q_1$  and  $q_3$  are of shifted period, and  $s(q_1, q_3) \equiv s(q_1, q_2) + s(q_2, q_3) \pmod{|w|}$ .

If  $q_1$  is of period  $w$ , then  $q_2$  is of period  $w_2 = w' w w'^{-1}$ , where  $w'$  is the suffix of  $w$  of size  $s(q_1, q_2)$ . If  $q_2$  is of period  $w_2$ , then  $q_2$  is of period  $w_3 = w'' w_2 w''^{-1}$ , where  $w''$  is the suffix of  $w_2$  of size  $s(q_2, q_3)$ . We then have that  $w_3 = (w'' w') w (w'' w')^{-1}$ , where  $(w'' w')$  is of size  $s(q_1, q_2) + s(q_2, q_3)$ .

We can compute the shift of the period of each state accessible from  $q$  rule by rule using the additive property of the shifts we proved above.  $\square$

## C Proof of Lemma 3

**Lemma.** *Let  $M$  be an LTW and  $q$  be a state in  $M$  that is quasi-periodic on the left. Let  $M^q$  be constructed by Algorithm 1 and  $p^e$  be a state in  $M^q$  accessible from  $q^e$ . Then  $M$  and  $M^q$  are equivalent and  $p^e$  is earliest.*

*Proof.* To show that  $M$  and  $M^q$  are equivalent we show that  $\text{lcp}(q) \llbracket M^q \rrbracket_{q^e}(t) = \llbracket M \rrbracket_q(t)$ , for all  $t \in \text{dom}(q)$ . To show that  $\text{lcp}(q) \llbracket M^q \rrbracket_{q^e}(t) = \llbracket M \rrbracket_q(t)$  we show that, for all states  $p$  accessible from  $q$  and all  $t \in \text{dom}(p)$ ,  $\llbracket M^q \rrbracket_{p^e}(t)$  and  $\rho_{s(q,p)} [\text{lcp}(p)^{-1} \llbracket M \rrbracket_p(t)]$  are equivalent as then

$$\begin{aligned} \text{lcp}(q) \llbracket M^q \rrbracket_{q^e}(t) &= \text{lcp}(q) \rho_{s(q,q)} [\text{lcp}(q)^{-1} \llbracket M \rrbracket_q(t)] \\ &= \text{lcp}(q) \text{lcp}(q)^{-1} \llbracket M \rrbracket_q(t) \\ &= \llbracket M \rrbracket_q(t). \end{aligned}$$

To show that  $\llbracket M^q \rrbracket_{p^e}(t) = \rho_{s(q,p)} [\text{lcp}(p)^{-1} \llbracket M \rrbracket_p(t)]$  for all  $p$  accessible from  $q$  and all  $t = (s_1, \dots, s_n) \in \text{dom}(p)$ , we prove that  $\llbracket M^q \rrbracket_{p^e}(t)$  is of the same period and of the same size as  $\rho_{s(q,p)} [\text{lcp}(p)^{-1} \llbracket M \rrbracket_p(t)]$ . From Lemma 2 we know that all  $p$  accessible from  $q$  are quasi-periodic and therefore  $\text{lcp}(p)^{-1} \llbracket M \rrbracket_p(t)$  is periodic. Hence, if  $\llbracket M^q \rrbracket_{p^e}(t)$  and  $\rho_{s(q,p)} [\text{lcp}(p)^{-1} \llbracket M \rrbracket_p(t)]$  are of the same period and of the same size then they are equivalent.

To show that  $\llbracket M^q \rrbracket_{p^e}(t)$  and  $\rho_{s(q,p)} [\text{lcp}(p)^{-1} \llbracket M \rrbracket_p(t)]$  have the same size, for all  $t \in \text{dom}(p)$ , we show that  $|\llbracket M^q \rrbracket_{p^e}(t)| = |\llbracket M \rrbracket_p(t)| - |\text{lcp}(p)|$ . The proof is by



induction on the input tree. For an input tree  $t$  with no subtrees we have

$$\begin{aligned} |\llbracket M^q \rrbracket_{p^e}(t)| &= |\rho_{s(q,p)}[\text{lcp}(p)^{-1}u_0 \dots u_n]| \\ &= |u_0| + \dots + |u_n| - |\text{lcp}(p)| \\ &= |\llbracket M \rrbracket_p(t)| - |\text{lcp}(p)| \\ &= |\rho_{s(q,p)}[\text{lcp}(p)^{-1}\llbracket M \rrbracket_p(t)]|. \end{aligned}$$

Thus, the base case holds. Consider an input tree  $t = f(s_1, \dots, s_n) \in \text{dom}(p)$ . Then  $\llbracket M^q \rrbracket_{p^e}(t)$  is of size  $|u_p| + |\llbracket M^q \rrbracket_{q_1^e}(s_{\sigma(1)})| + \dots + |\llbracket M^q \rrbracket_{q_n^e}(s_{\sigma(n)})|$  with  $|u_p| = |\rho_{s(q,p)}[\text{lcp}(p)^{-1}u_0 \text{lcp}(q_1) \dots \text{lcp}(q_n)u_n]|$ . Since shifting a word preserves its length, we have  $|u_p| = |u_0| + |\text{lcp}(q_1)| + \dots + |u_n| - |\text{lcp}(p)|$ . Thus, we have to show that

$$\begin{aligned} |u_p| + |\llbracket M^q \rrbracket_{q_1^e}(s_{\sigma(1)})| + \dots + |\llbracket M^q \rrbracket_{q_n^e}(s_{\sigma(n)})| &= |\rho_{s(q,p)}[\text{lcp}(p)^{-1}\llbracket M \rrbracket_p(t)]|. \\ \text{By induction we have } |\llbracket M^q \rrbracket_{q_i^e}(s_{\sigma(i)})| &= |\llbracket M \rrbracket_{q_i}(s_{\sigma(i)})| - |\text{lcp}(q_i)|. \text{ Thus, we have} \\ |u_p| + |\llbracket M^q \rrbracket_{q_1^e}(s_{\sigma(1)})| + \dots + |\llbracket M^q \rrbracket_{q_n^e}(s_{\sigma(n)})| \\ &= |u_0| + |\text{lcp}(q_1)| + \dots + |\text{lcp}(q_n)| + |u_n| - |\text{lcp}(p)| \\ &\quad + |\llbracket M \rrbracket_{q_1}(s_{\sigma(1)})| - |\text{lcp}(q_1)| + \dots + |\llbracket M \rrbracket_{q_n}(s_{\sigma(n)})| - |\text{lcp}(q_n)| \\ &= |u_0| + |\llbracket M \rrbracket_{q_1}(s_{\sigma(1)})| + \dots + |\llbracket M \rrbracket_{q_n}(s_{\sigma(n)})| + |u_n| - |\text{lcp}(p)| \\ &= |\llbracket M \rrbracket_p(t)| - |\text{lcp}(p)| \\ &= |\rho_{s(q,p)}[\text{lcp}(p)^{-1}\llbracket M \rrbracket_p(t)]|. \end{aligned}$$

To show that  $\llbracket M^q \rrbracket_{p^e}(t)$  and  $\rho_{s(q,p)}[\text{lcp}(p)^{-1}\llbracket M \rrbracket_p(t)]$  have the same period, for all  $t \in \text{dom}(p)$ , we show that  $\llbracket M^q \rrbracket_{p^e}(t) \in u^*$  and  $\rho_{s(q,p)}[\text{lcp}(p)^{-1}\llbracket M \rrbracket_p(t)] \in u^*$  where  $L_q \subseteq uu^*$ . From Lemma 2 it follows that  $\rho_{s(q,p)}[\text{lcp}(p)^{-1}\llbracket M \rrbracket_p(t)] \in u^*$ . To proof that  $\llbracket M^q \rrbracket_{p^e}(t) \in u^*$  is by induction on the input tree. For an input tree  $t$  with no subtrees we have  $\llbracket M^q \rrbracket_{p^e}(t) = \rho_{s(q,p)}[\text{lcp}(p)^{-1}u_0 \dots u_n]$ . From Lemma 2 we know that  $L_p$  is quasi-periodic of period  $u'u''$  where  $u = u''u'$  and  $u'$  is of size  $s(q,p)$ . Thus,  $\text{lcp}(p)^{-1}u_0 \dots u_n \in (u'u'')^*$  and therefore  $\rho_{s(q,p)}[\text{lcp}(p)^{-1}u_0 \dots u_n] \in (u'u'')^* = u^*$ . Hence, the base case holds. Consider an input tree  $t = f(s_1, \dots, s_n) \in \text{dom}(p)$ . Then we have  $\llbracket M^q \rrbracket_{p^e}(t) = \rho_{s(q,p)}[\text{lcp}(p)^{-1}u_0 \text{lcp}(q_1) \dots \text{lcp}(q_n)u_n] \llbracket M^q \rrbracket_{q_1^e}(s_{\sigma(1)}) \dots \llbracket M^q \rrbracket_{q_n^e}(s_{\sigma(n)})$ . By induction,  $\llbracket M^q \rrbracket_{q_i^e}(s_{\sigma(i)}) \in u^*$ . With the same argumentation as in the base case  $\text{lcp}(p)^{-1}u_0 \text{lcp}(q_1) \dots \text{lcp}(q_n)u_n \in (u'u'')^*$  with  $u = u''u'$  and  $u'$  is of size  $s(q,p)$ . Thus,  $\rho_{s(q,p)}[\text{lcp}(p)^{-1}u_0 \text{lcp}(q_1) \dots \text{lcp}(q_n)u_n] \in (u'u'')^* = u^*$  and therefore we get  $\llbracket M^q \rrbracket_{p^e}(t) \in u^*$ .  $\square$

## D Proof of Lemma 4

**Lemma.** *Let  $q$  be a state of an LTW  $M$  and  $T^q$  be constructed by Algorithm 2. Then  $M$  and  $T^q$  are same-ordered and  $q$  is quasi-periodic on the left if and only if  $\llbracket M \rrbracket_q = \llbracket T^q \rrbracket$  and  $q^e$  is periodic.*

*Proof.* We show that  $q$  is quasi-periodic on the left if and only if  $\llbracket M \rrbracket_q = \llbracket T^q \rrbracket$  and  $q^e$  is periodic. If  $q$  is quasi-periodic on the left the transformation in Algorithm 2 is the same as in Algorithm 1. Therefore  $\llbracket M \rrbracket_q = \llbracket T^q \rrbracket$  and  $q^e$  is periodic.

If  $\llbracket M \rrbracket_q = \llbracket T^q \rrbracket$  and  $q^e$  is periodic, then  $\llbracket M \rrbracket_q$  is quasi-periodic as  $\llbracket T^q \rrbracket = w_q \llbracket T^q \rrbracket_{q^e}$  with  $w_q$  a shortest word of  $L_q$ .

$M$  and  $T^q$  are same-ordered as the order of the rules in  $T^q$  is the same as in  $M$  by construction.  $\square$

## E Proof of Theorem 5

**Theorem.** *Let  $M$  be an LTW. Then an equivalent LTW  $M'$  where all quasi-periodic states are earliest can be constructed in polynomial time.*

*Proof.* This proof works by induction. We first show that if  $M = (\Sigma, \Delta, Q, \text{ax}, \delta)$  has  $n$ ,  $n \geq 1$  quasi-periodic states that are non-earliest, then we can build in polynomial time an equivalent LTW  $M'$  with  $n - 1$  non-earliest quasi-periodic states. Using Algorithm 2 we choose  $q$  as a non-earliest quasi-periodic state of  $Q$ . We apply Algorithm 1 on state  $q$  and get  $M^q$ , whose set of state is of form  $Q \sqcup Q^e \setminus \{q\}$ , where  $Q^e$  is the set of states  $p^e$  with  $p$  accessible from  $q$  that are created by Algorithm 1. According to Lemma 3 all the states of  $Q^e$  are periodic. This means that the non-earliest quasi-periodic states of  $M^q$  are all in  $Q \setminus \{q\}$ . Since  $Q$  has  $n$  non-earliest quasi-periodic states, including  $q$ ,  $M^q$  has  $n - 1$ .

Now we can build  $M_1$  equivalent to  $M$  with  $n - 1$  non-earliest quasi-periodic states, then  $M_2$  equivalent to  $M_1$  (hence to  $M$ ) with  $n - 2$  non-earliest quasi-periodic states, and so on. Finally we get  $M_n$  equivalent to  $M$  with no non-earliest quasi-periodic state. Each step is in polynomial time and the number  $n$  is smaller than the number of states in  $M$ . For each occurrence of a state on the right-hand side of a rule there is at most one new state needed in the construction. Therefore the size increase of the transducer is only polynomial. To avoid the construction of equivalent states  $q$  should be considered before  $q'$  if  $q'$  is accessible from  $q$ . If  $q$  is accessible from  $q'$  and  $q'$  is accessible from  $q$  then  $q$  is considered first if there is a acyclic way from the axiom to  $q$  that contains  $q'$ .

In the above proof we assumed that Algorithm 2 and 1 run in polynomial time. In both algorithms it is crucial that SLPs are used to represent the shortest words of the languages produced by the states of a transducer as these can be of exponential size, cf. Example 1. Instead of these uncompressed words nonterminals representing these words as SLPs are inserted in the transducers. All operations that are needed in the algorithms, namely constructing a SLP for the shortest word of an CFG, concatenation of SLPs, shifting the word produced by an SLP and removing the prefix or suffix of an SLP are in polynomial time [11].  $\square$

## F Proof of Theorem 6

**Theorem.** *For each LTW  $M$  where all quasi-periodic states are earliest we can build in polynomial time an equivalent LTW  $M'$  such that each part  $q(x)u$  of a*

rule in  $M$  where  $L_q u$  is quasi-periodic is earliest.

*Proof.* First, we show that one part  $q(x)u$  of a rule where  $L_q u$  is quasi-periodic on the left can be replaced by their earliest form. We can apply Algorithm 2 and 1 on a part  $q(x)u$  of a rule that is quasi-periodic by replacing  $q(x)u$  by a new state  $\hat{q}$ . Therefore each occurrence of  $q(x)u$  in any rule is replaced by  $\hat{q}(x)$  and for each rule  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$  a rule  $\hat{q}, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n u$  is added. Then we can apply Algorithm 2 on  $\hat{q}$  to test  $q(x)u$  for quasi-periodicity on the left. If  $\hat{q}u$  is quasi-periodic on the left we apply Algorithm 1 on  $\hat{q}$  to replace  $q(x)u$  by their earliest form.

Second, we show that for any rule of an LTW an equivalent rule can be constructed such that all parts  $q(x)u$  in the rule where  $L_q u$  is quasi-periodic on the left are earliest. Consider a rule  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$ . Replacing a part  $q(x)u$  with  $L_q u$  is quasi-periodic by their earliest form as described above means that all occurrences of  $\hat{q}(x)$  are replaced by  $\text{lcp}(\hat{q})\hat{q}^e(x)$ . Thus, to replace all parts of a rule that produce quasi-periodic languages by their earliest form the testing and replacing should be done from right to left as the earliest form may introduce new words on the left of the replaced state. For a rule  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$  we start by testing  $q_n(x_{\sigma(n)}) u_n$  for quasi-periodicity on the left and replace the part if necessary as described above. If so, we obtain  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_{n-1}(x_{\sigma(n-1)}) u_{n-1} \text{lcp}(\hat{q}_n) \hat{q}_n^e(x_{\sigma(n)})$  and continue with testing  $q_{n-1}(x_{\sigma(n-1)}) u_{n-1} \text{lcp}(\hat{q}_n)$  for quasi-periodicity on the left. If not, then we continue with testing  $q_{n-1}(x_{\sigma(n-1)}) u_{n-1}$  for quasi-periodicity on the left. Following this construction for the rule from right to left, i.e. from index  $n$  to 1, leads to an equivalent rule where all parts  $q(x)u$  with  $L_q u$  is quasi-periodic on the left are earliest.

The construction runs in polynomial time as Algorithm 2 and 1 run in polynomial time (for details see the proof of Theorem 5) and for each rule  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$  the algorithms are applied at most  $n$  times.  $\square$

## G Proof of Lemma 6

**Lemma.** *If  $M$  and  $M'$  are equivalent and in partial normal form then they are same-ordered.*

*Proof.* The proof is by contradiction. Suppose that  $M, M'$  are two equivalent LTWs in partial normal form that are not same-ordered. We consider co-reachable state  $q$  and  $q'$  of  $M$  and  $M'$  that are not same-ordered. Then for  $q, f \rightarrow u_0 q_1(x_{\sigma(1)}) \dots q_n(x_{\sigma(n)}) u_n$  and  $q', f \rightarrow v_0 q'_1(x_{\sigma'(1)}) \dots q'_n(x_{\sigma'(n)}) v_n$  we choose  $i, j, i < j$  such that  $j - i$  is minimal under the following constraints

- $\{\sigma(k) \mid i \leq k \leq j\} = \{\sigma'(k) \mid i \leq k \leq j\}$  and
- there is  $k, i \leq k \leq j$  such that  $\sigma(k) \neq \sigma'(k)$ .

As the set of indices of  $\sigma(k)$  and  $\sigma'(k)$  between  $i$  and  $j$  is the same but in different orders there is  $l, i \leq l \leq j$  such that  $\sigma(l) > \sigma(l+1)$  or  $\sigma'(l) > \sigma'(l+1)$ . W.l.o.g. we assume that  $\sigma(l) > \sigma(l+1)$ . Additionally, from the above constraints it

follows that there is  $s, t, i \leq s \leq l < t \leq j$  such that  $\sigma(s) = \sigma'(t)$ . As the partial normal form satisfies the preconditions of Theorem 4 we get that  $q_s, \dots, q_t$  are periodic with the same period and  $u_s, \dots, u_{t-1}$  are empty. Thus,  $L_{q_t} u_t L_{q_{t+1}}$  is quasi-periodic. Then it follows from the partial normal form that  $\sigma(l) < \sigma(l+1)$ , a contradiction. □

## H Proof of Theorem 8

**Theorem.** *The equivalence problem for linear tree-to-word transducers is decidable in polynomial time.*

*Proof.* Let  $M$  and  $M'$  be two linear tree-to-word transducers. We construct equivalent LTWS  $M_1$  and  $M'_1$ , respectively, such that  $M_1$  and  $M_2$  are in partial normal form following Theorem 7. We then test if  $M_1$  and  $M'_1$  are same-ordered. If they are same-ordered we test  $M_1$  and  $M'_1$  for equivalence, see Theorem 1. If  $M_1$  and  $M'_1$  are not same-ordered we know following Lemma 6 that  $M_1$  and  $M'_1$  are not equivalent and therefore  $M$  and  $M'$  are not equivalent. □